Department of the Air Force Chief Data Office Overview

Leveraging Data as a Strategic Asset

Ms. Eileen Vidrine
Chief Data Officer
US Air Force & US Space Force

SAF/CO Overview

- **Mission: To empower and posture the Department of the Air Force (DAF) to harness data for competitive military advantage**
- **Reports directly to Under Secretary of the Department of the Air Force (USecAF)**
- **Established in September of 2017**
- **Program Guidance Letter signed on 19 March 2019**
- **Mission Directive signed on 30 April 2019**
- **Full Operational Capability achieved 1 June 2020**

"Next-generation combat is going to depend on data, more so than any time in the past. It's important for us to make sure all of our Airmen have uninterrupted access to all of the data they need, when they need it."

– Under Secretary of the Air Force Matthew Donovan

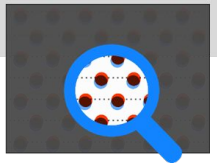
SAF/CO Problem Space

- **Large number of legacy systems with numerous point-to-point interfaces that are expensive to implement and maintain**
- **Antiquated, brittle architectures inhibit legacy system enhancements required to make data accessible**
- **Many legacy systems employ "data jails" that prevent timely data access that supports the analytics needs of Functional communities**
- **The time is now to employ new, commercially available data services that will change the landscape of how data enables mission effectiveness**

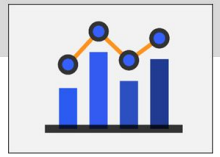
Source: CDO AFDAWG Kickoff Brief, June 2018

Data Challenges in Air Force Today

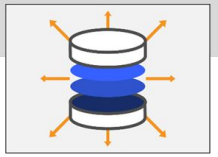
Lack Visibility
Unable to assess health and readiness because of limited access to data.



Limited Analytics Capability
Cannot perform robust analytics to make mission decisions.



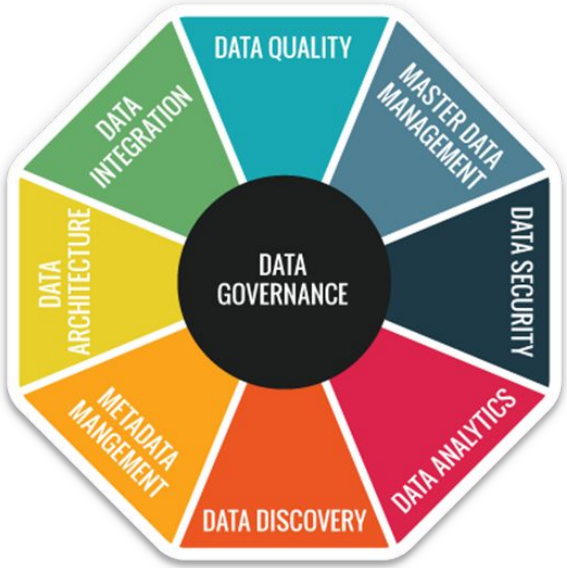
Growing Problem
Generate more data every second and keep holding data hostage in data jails.



Source: CDO brief to USECAF, January 2019

The DAF CDO sees the challenges in the context of the fully functional scope of Enterprise Data Management disciplines and achieving targeted degrees of EDM Maturity

Out of which Data Analytics is one piece of the overall problem space.



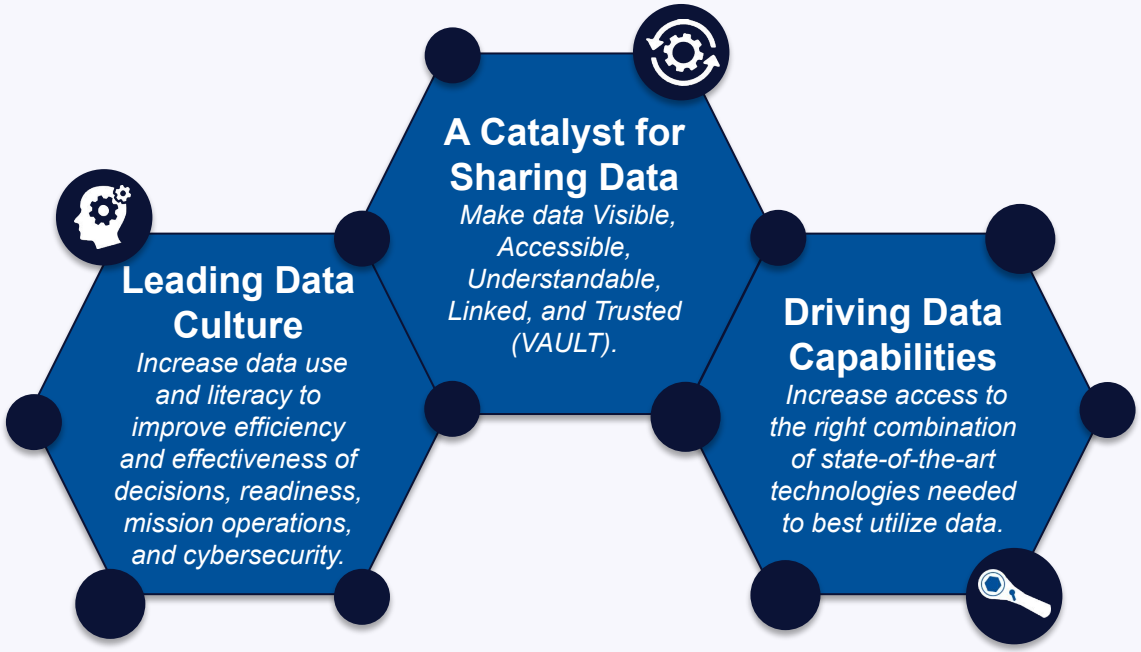
SAF/CO Strategic Vector

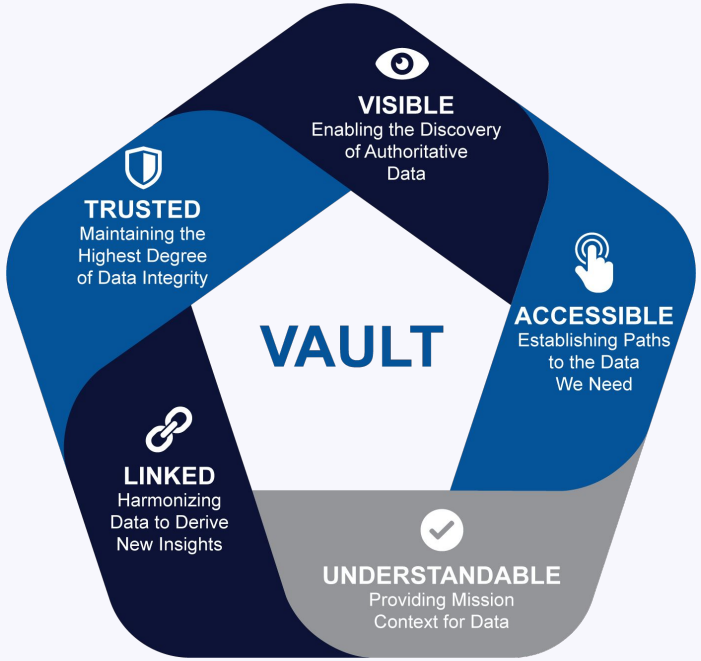
VISION & MISSION

VISION - ***Fueling data innovation within the Department of the Air Force***

MISSION - ***To empower and posture the Department of the Air Force to harness data for competitive military advantage***

STRATEGIC GOALS





Digital Air Force Journey

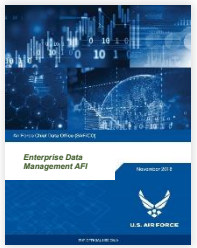
Overarching Digital AF Journey

Worldwide Mission Readiness & Effectiveness

DAF Data Transformation

DAF Data Reference Architecture:
Leads DAF organizations toward implementing Data Platforms that are able to share data

VAULT Data Platform and Enterprise Information Model:
Provides cyber-secure, cloud-based tools to connect, find, share, and learn from DAF data



Enterprise Data Management AFI:
Defines guidance, roles, and responsibilities that enable an DAF culture that treats data as a strategic asset

SAF/CO is driving strategic and tactical actions, moving the DAF toward being a data-driven organization prepared for Digital Transformation

VAULT Data Platform Lines of Effort

LOE 1: Data Services and Tools

Cloud-hosted capabilities for DAF mission areas to identify, catalog, describe, expose, cleanse, transmit, collect, process, manage, analyze, and harvest business value from data across the enterprise.

LOE Accomplishments

- ✔ NIPR and SIPR Environments
- ✔ Visualization w Tableau
- ✔ Data Ingest Service
- ✔ Enterprise Info Mgmt Service
- ✔ Engage community to populate EIM
- ✔ AI/ML using SPARK in Databricks
- ✔ Data Cleansing Pipelines with Trifacta
- ❑ Consumption-based business model
- ❑ Many more on Roadmap

LOE 2: Data Indexing and Exposure

Make current and trusted data sources available for analytics across the enterprise. Critical DAF questions (e.g. Readiness) are dependent upon having fast and easy visibility, access and understanding of data from many functional data areas.

LOE Accomplishments

- ✔ Secure and curate EXORD data sets
- ✔ Publish user access processes
- ✔ Actively manage the environment
- ✔ Automate data updates
- ✔ COVID19 data from Advana
- ✔ Financial data from DFAS
- ❑ Handle Unstructured & Streaming Data
- ❑ Expand to more sources

LOE 3: Analytics Partnerships

Enable functional communities to use the analytical tools they need. The VAULT Data Platform is vendor and tool agnostic. Any products that adhere to security requirements and the DAF Data Reference Architecture can be used in the VAULT Data Platform.

LOE Accomplishments

- ✔ AFCEC Visualizations
- ✔ Aircrew Tasking Custom App
- ✔ Custom A9 R-Shiny Applications
- ✔ A3 Readiness Dashboard NIPR/SIPR
- ✔ 50+ NIPR and SIPR Tenants in F35, MROi, AFMAA, AFCAC, many more
- ✔ Data exchanges with DoD's Advana
- ❑ Support ABMS / JADC2

The VAULT Journey

NIPR VAULT

July 2018
- VAULT Contract Award

Oct 2018
- NIPR IL-4 ATO
- OASIS Readiness Task

Jan 2019
- NIPR IL-5 ATO
- 10TB static data

Apr 2019
- Tableau Go-Live
- 200 Global NIPR Users
- 3 Tenant Projects

Oct 2019
- Databricks NIPR Go-Live
- SwimLane (SBIR) Go-Live
- 10 NIPR Tenant Projects
- 1 live data feed

Apr 2020
- COVID-19 Reports
- User Account Creation Automated
- R-Shiny App Server
- 25 Tenant Projects
- 5 live data feeds

Aug 2020
- Over 40 Tenant Projects
- 50k DBUs added
- 2000+ Users
- Trifacta and NNCompass Go-Live
- 12 live data feeds, including Advana

SIPR VAULT

Aug 2019
- SIPR VAULT Contract Award

Mar 2020
- SIPR IL-6 ATO
- 15 SIPR Users
- 1 Tenant Project (Readiness Pathfinder)
- AWS Diode in use

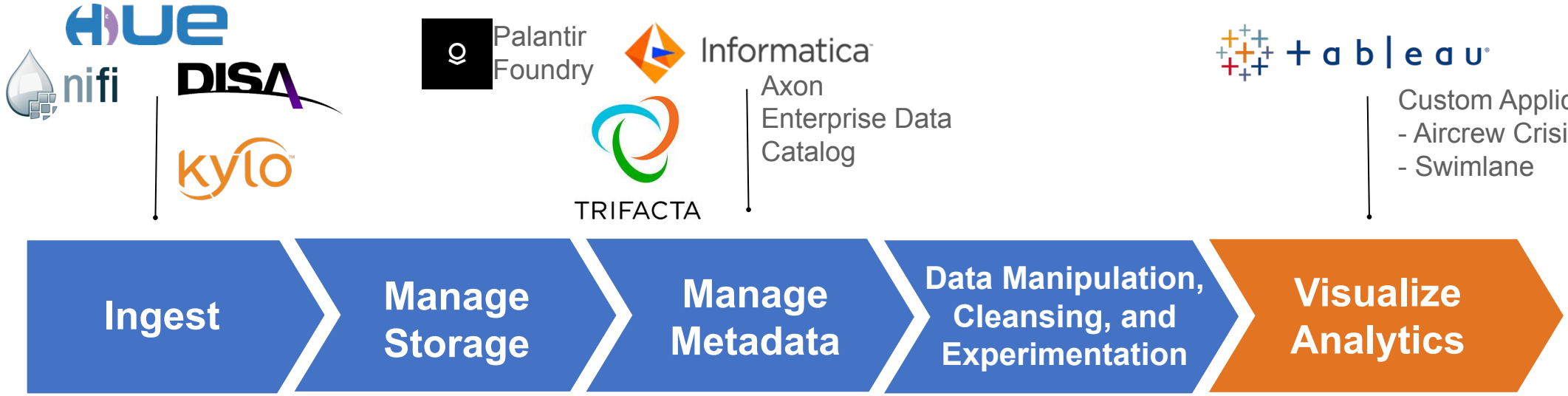
June 2020
- R-Shiny Go-Live
- COVID-19 Project
- 3 SIPR Tenant Projects

Aug 2020
- 8 SIPR Tenant Projects
- Databricks SIPR Go-Live
- 90 Users
- Trifacta and NNCompass being deployed

Progress on all SAF/CO Lines of Effort

- **New Open Source and COTS tools deployed**
- **New data sets acquired**
- **50 Analytics Partnerships between NIPR and SIPR, over half of which < 90 days old**
- **User growth from 200 to 2000 in 1 year**
- **Tableau views from 16 to 1,800 in 1 year**
- **Same number of developers and Operations & Support staff as Oct 2018**

VAULT Platform Capabilities Today

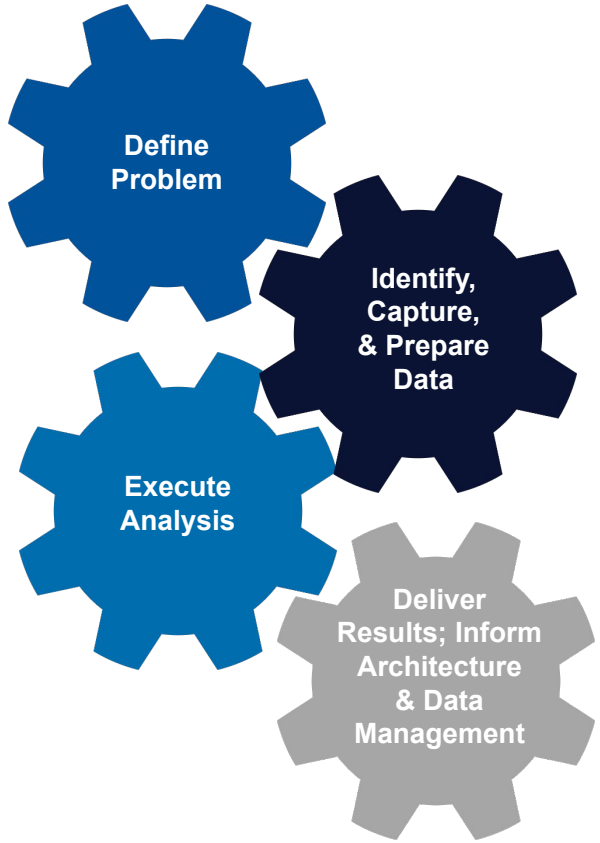


Features:

- **ATO: IL-4, 5, & 6, PII**
- **Multi-Tenant**
- **Group/Role Access on Data and Apps**
- **DISA CAP Connection**
- **User-unique Webtop**

SAF/CO Use Cases

- Define mission-related problems that can be addressed with data solutions
- Capture and prepare data
- Optimize analysis through custom ML or leveraging existing models
- Glean insights through visualizations



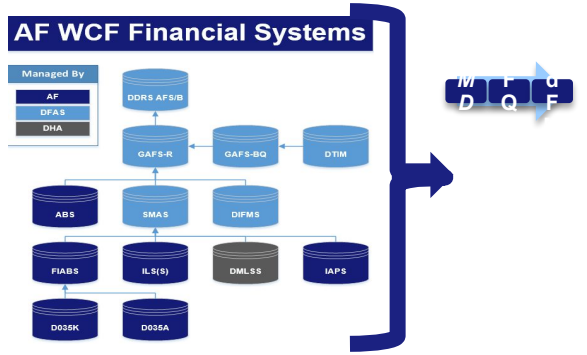
SAF/CO's Use Case-driven approach leverages data and applied analytics to solve problems and drive mission innovation across the DAF

AFMC/FM Data Analytics

CDO VAULT Utilization

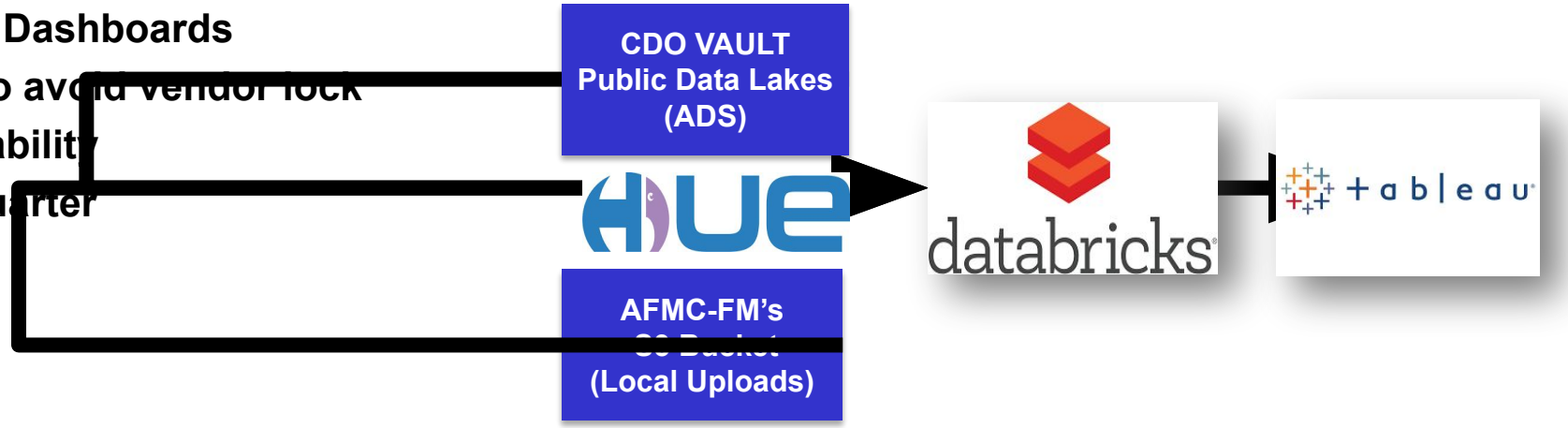
Purpose

- Provide insight into the WCF Cash balances by being able to:
  - Provide insight into underlying business events
  - Forecasting
  - Visualize the impact of varying operating scenarios



- Databricks
  - Hub for Extract, Transform, and Loading (ETL) functions
  - Direct data refresh of Tableau Dashboards
  - Allows us to code in Python to avoid vendor lock
  - Strong performance and scalability
  - Over 65 million records per quarter

AFMC/FM Current "Critical Software Path"



AFMC/FM Data Analytics

AFMC Cash Reporting Dashboard

Purpose

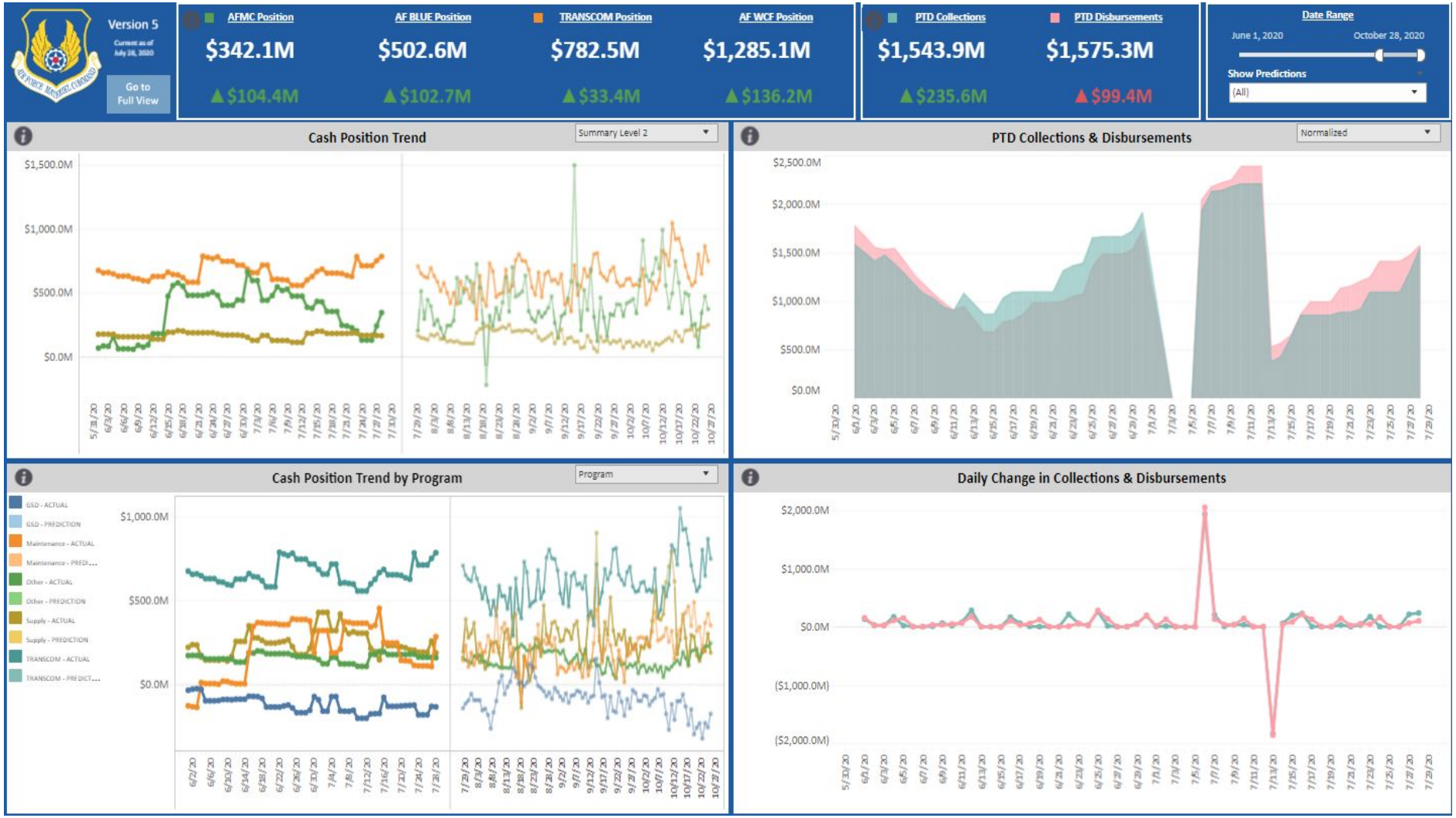
- Provide insight into underlying business events
- Forecasting
- Visualize the impact of various operating scenarios

Data Sources

- DFAS ESS Produced Daily Flash Cash (local upload)
- GAFS-BQ STH (ADS)

Operations

- Automated - STH ingestion, ETL Efforts (both sources), Dashboard data refresh
- Manual – Flash Cash ingestion



AFMC/FM Data Analytics

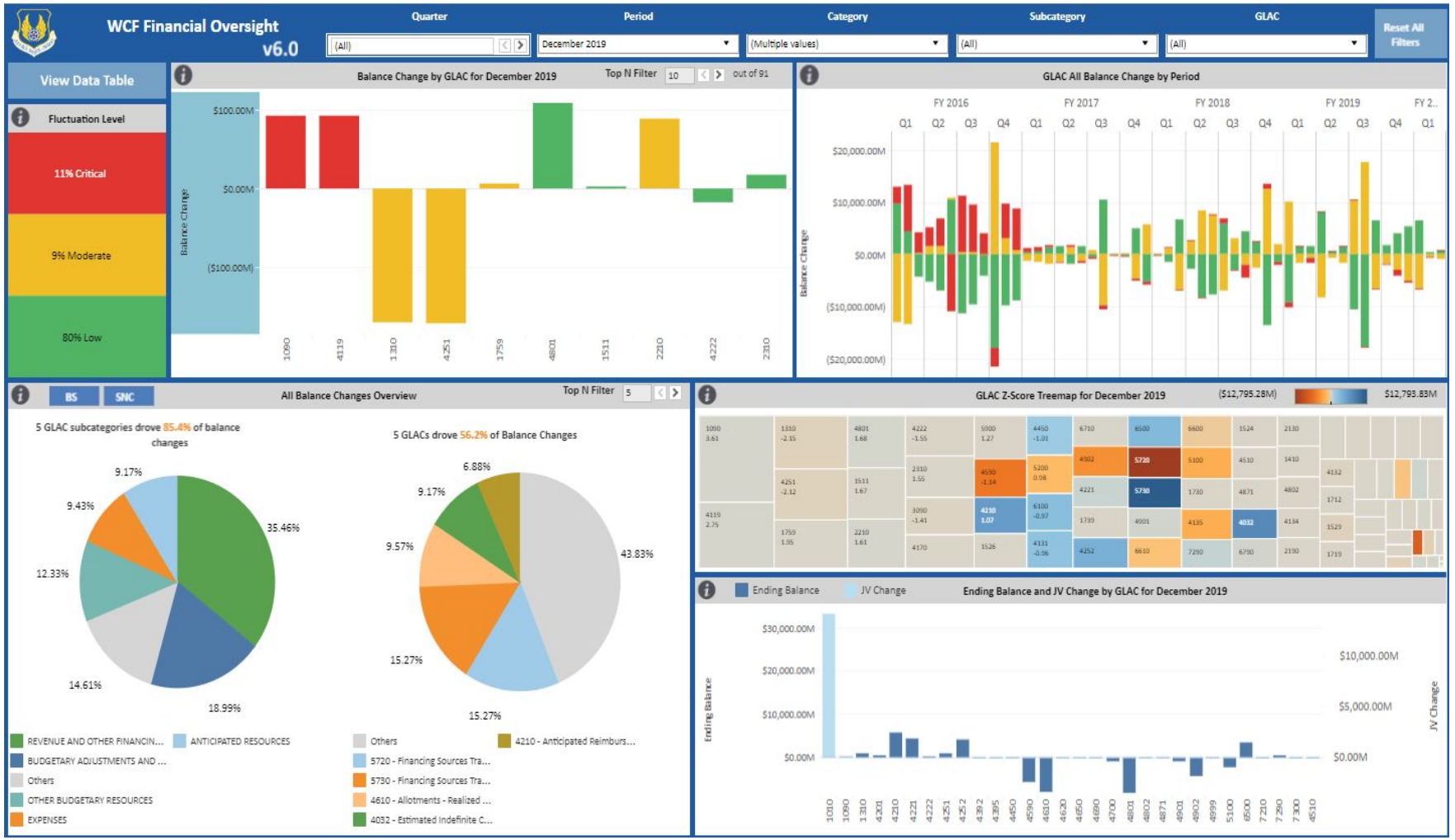
WCF FM Oversight

Purpose

- Enable the WCF FM community to:
  - Engage in Financial Statement analysis
  - Link Financial Statement balances to WCF GL Transactions
  - Detect Abnormal Balances or Balance Changes

Data Sources

- GAFS-R Trial Balances and Journal Vouchers
- SMAS, FIABS, DIFMS, and ABS



Really cool versus Really Awesome

Initially - DAF Data Analytic Pipeline Landscape

Lots of 'Master Builders'

Where we maturing to……..

Architected: alignment & clarity

Thank you!

BACKUP

VAULT Data Platform
Service and Software Overview

Platform Tenant Instance Software/Applications/Services	NIPR	SIPR
Apache Zeppelin	●	●
Apache HUE	●	●
Apache Kylo	●	●
Apache NiFi	●	●
Apache NiFi UI	○	○
Informatica Axon (Read Only)	●	●
Apache ATLAS	●	○
Apache Spark	●	●
AWS Elastic MapReduce	●	●
AWS Athena	●	○
AWS Sagemaker	●	○
Tableau Server ($)	●	●
Tableau Desktop ($)	●	○
Databricks ($)	●	●
Swimlane ($)	●	○
Trifacta ($)	●	◑
NNData/NNCompass ($)	◕	◑
AWS S3 Browse	●	●
ServiceNow SaaS ITOM ($)	●	○
ServiceNow SaaS ITSM ($)	●	○
ServiceNow SaaS ITBM ($)	●	○
ServiceNow SaaS SAM ($)	●	○
ACTF App	●	N/A
AFBOP ServiceNow On-Prem ($)	N/A	○
Kibana	●	○

Platform Software Languages / Libraries	NIPR	SIPR
Scala	●	●
R	●	●
Python	●	●
PL/SQL	●	●
OpenJDK/Java	●	●
OpenJDK	○	○
NodeJS	●	○
PyTorch	●	○
Anaconda 3	○	○
Bokeh	●	○
Colorset	●	○
Behave	●	○
Flask SQL Alchemy	○	○
Swagger API	●	○
Gunicorn	●	○
Tensorflow	○	○
MLFlow	○	○
Plotly Dash	○	○

Platform Infrastructure Compute Technologies	NIPR	SIPR
Redhat	●	●
Microsoft Windows	●	●
AWS EC2	●	●
AWS CLI	●	●
Shiny Server	●	●
Nginx	●	●
Apache HTTP Server	●	●
Apache Elasticsearch	●	●
Apache Kibana	●	●
Apache Tez	●	○
Apache Ambari	●	●
Apache Sqoop	●	●
Apache Kafka	●	●
Openshift	○	●
Kubernetes	◑	●
Docker	○	○
AWS Kinesis	●	●
Apache Solr	●	●

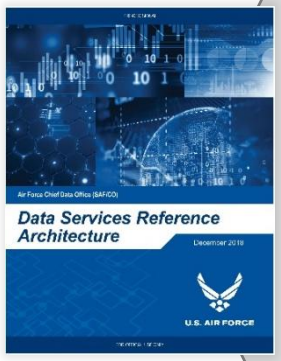
Platform Infrastructure Storage Technologies	NIPR	SIPR
AWS S3 Storage	●	●
AWS Relational DB Service	●	●
Apache Hadoop	●	●
Apache HBASE	●	●
S3-Fuse	●	●
Oracle RDBMS	●	●
PostgresSQL	●	●
MySQL	●	●
MariaDB	●	●
AWS SNS	●	○
AWS SQS	●	○
Apache Zookeeper	●	○

Addt'l Platform Technologies Security, Data Transfer, Release Mgmt	NIPR	SIPR
Big IP F5	●	●
Microsoft Active Directory	●	●
ACAS	●	●
Apache Knox	●	●
Apache Ranger	●	●
Splunk	●	●
HBSS	●	●
OpenSCAP CL Tool	●	●
SFTP	●	●
SMTP	●	●
SonarQube	○	○
Git	●	○
Jenkins	●	○
Cucumber	◑	●
Selenium	●	●

○ Roadmap Item ◑ Verifying (Tech & IA) Compliance ● Deployed / Available

◔ Prioritized by CDO Client ◕ Change Mgmt / CCB Review $ - Licensed SW

SAF/CO Data Services Reference Architecture



DAF Data Services Reference Architecture v1.0 has been signed by Mr. Donovan, Acting SECAF!!

DATA PRODUCT CONSUMER SERVICES

- Ad Hoc Reporting
- Delegate Remediation Facility
- Application Prototyping
- Enterprise Data Quality Dashboard
- Data Visualization
- Enterprise Metadata Pub/Sub Portal
- Root Cause Analysis Facility
- Standard Business Intelligence
- On Demand Data Retrieval Sandbox

Microservices:

Data Visualization | Data and Analytics Catalog Interface | Data Product and Service Pub/Sub Interface | Testing Tools | Virtual Project Workspace | Collaboration Spaces | Data Query Interface | Analytics Scripting Interface | Workflow Scripting Interface Reporting

ENTERPRISE DATA & ANALYTIC SERVICES

- Data Ingest
- ETL Load Object Accounting
- Metatagging
- Data Quality Assessment
- Data Profiling
- Business Transformation
- ETL Source-to-Target Mapping
- Data Quality Remediation

Microservices:

Data Flow Orchestration | Data Storage | Indexing & Search | Query Services | Analytical Processing Engine | Workflow Scripting Execution Engine | Analytics Libraries | Service Endpoints | Analytical Scripting Engine | Data Format Transformation

ENTERPRISE METADATA SERVICES (EIM)

- Data Modeling
- Metadata Model Administration
- Data Security Rules Management
- ADS Management
- Stewardship Metadata Administration and Metrics
- RBAC Roles and Credential Management
- Data Standards
- Business Rule Management

Microservices:

Business Rules Engine | Data Access Rules | Analytics Catalog | Service Publication Toolset | Services Registry | Role Based Access Control (Authorization) | Business Glossary | Analytics Content | Operational Metadata (Metrics) | Information Asset Catalog | Data Lineage | Enterprise Metadata Model

DATA PLATFORM FOUNDATION SERVICES

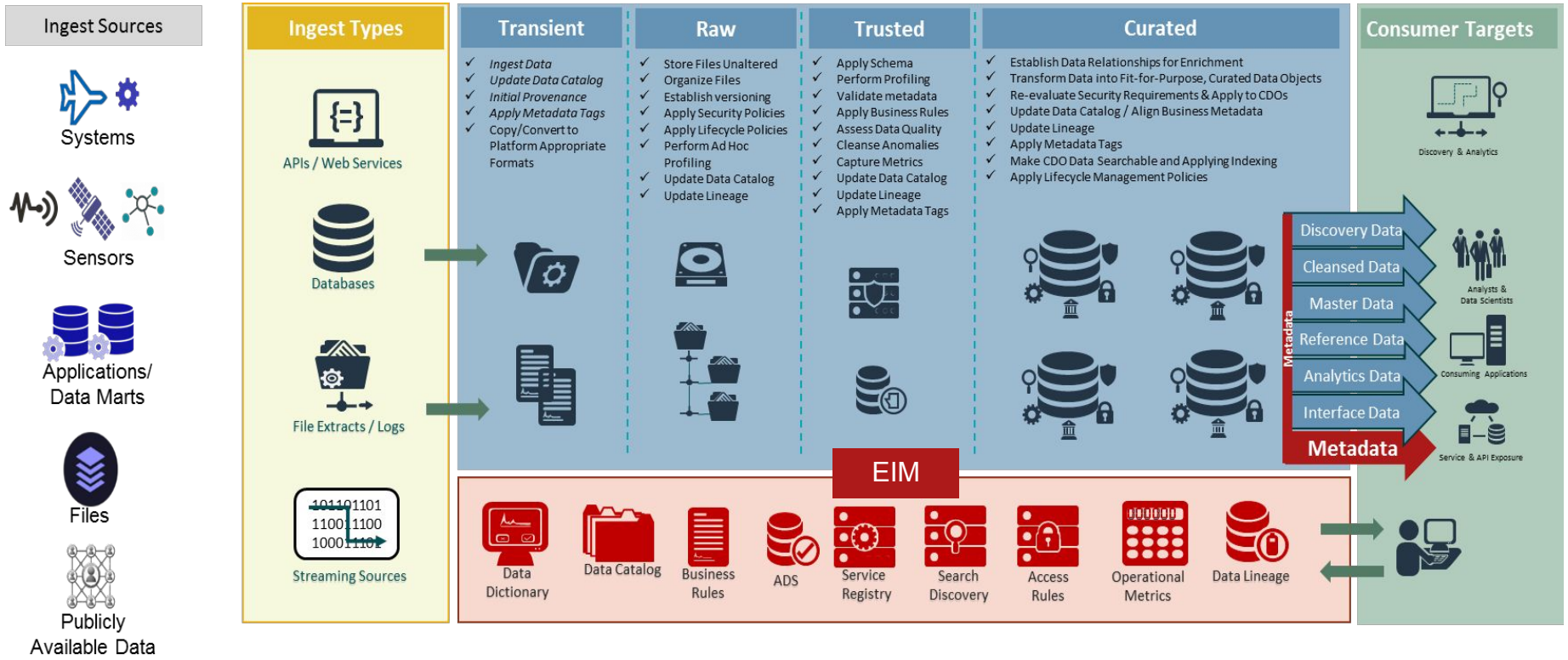
- Entity Administration Srvice
- Metadata Capture Automation (Pub/Sub)
- PGUID Generation, Administration, and Management
- Platform Global Runtime Substituion Variables
- On-Demand Deployment (Container Orchestration)

Microservices:

Instrumentation Services | Service Level Agreement Monitoring | Service Level Agreement Monitoring | Job Management | Entity Management | Global Data Service Activity Monitoring | Exception Management | System Interface Management | Message and Job Queuing

The Reference Architecture provides a blueprint for enabling consistent data access and usage across DAF. The VAULT Data Platform is fully compliant.

VAULT Data Platform Storage Concept and Organization



The DAF CDO Data Lake is structured and governed based on a Metadata Enabled (EIM) Data Curation Methodology